REPORT DOCUMENTATION PAGE

AFRL-SR-AR-TR-04-

0643

1. REPORT DATE *(DD-MM-YYYY)*
30-09-2004

2. REPORT TYPE
Final Report

3. DATES COVERED *(From - To)*
February, 2003 – June, 2004

4. TITLE AND SUBTITLE

Computational Methods for High Resolution Imaging and Data Mining

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA49620-03-1-0215

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)

Victor Paul Pauca – Principal Investigator

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Department of Computer Science
Wake Forest University
Box 7311
Winston-Salem, NC 27109

8. PERFORMING ORGANIZATION REPORT NUMBER

1

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Computational Mathematics
Air Force Office of Scientific Research / NM
4015 Wilson Blvd, Room 713
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release: distribution unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Air Force position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
Research problems investigated during the funding period included enabling mathematics and computational research in integrated optical systems design as well as spectral data mining and analysis. Particular applications of this research to enhance image restoration of integrated optical-digital systems and space object identification were addressed. A novel minimization-maximization approach was proposed and successfully employed to optimize the design of integrated imaging systems that maximize the information strength in an incoming image signal, before it is discretized and sampled in a CCD array. Five research papers were published on this work. Novel constrained nonnegative matrix factorization (CNMF) algorithms were developed and successfully used to extract material features and material composition from spectral traces of man-made space objects, such as geosynchronous satellites. A total of four papers were published as a result of this work. Technology transfer included collaboration with researchers from Oceanit Labs and Boeing Inc. (2 research papers) and transition of our CNMF algorithms for text mining and document clustering applications (2 research papers). Research results were presented at the 2004 SIAM International Conference on Data Mining, at the Annual SPIE meeting and at the AMOS Technical Conference in both 2003 and 2004.

15. SUBJECT TERMS
pupil-phase engineering, constrained nonnegative matrix factorizations, spectral data analysis, space object identification

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL		

FINAL REPORT DOCUMENTATION PAGE (SF298)
(Continuation Sheets - Grant: FA49620-03-1-0215, PI - V. Paúl Pauca)

1 Accomplishments/New Findings

This research project was focused on the enabling mathematics and computational research for two challenging research areas: a) integrated optical-digital system design for extended depth of field, and b) numerical algorithms for space object identification. Both research areas involve the formulation and numerical solution of difficult nonlinear optimization problems as well as the development of complex simulation systems.

Important accomplishments were made in both research areas. In the former novel optimization approaches were developed in order to improve the design of optical-digital systems. The numerical solution of these problems is not trivial and required significant processing power (approximately 1.2 CPU years). The results obtained include the discovery of new optical elements for extending the depth of field (portion of an image that is in focus) for which patents have been recently filed. Five papers [2,3,7,8,9] were published in this area during the reporting period, including two invited journal papers and three conference papers.

In the latter area, approaches that exploit novel constrained nonnegative matrix factorization methods (developed by the PI) as well as information theoretic measures were proposed for better identification and extraction of intrinsic "parts" or features from remote sensing data, for space object identification (SOI). The efficacy and accuracy of these methods were studied using simulated spectral data provided by Maile Giffin from Oceanit Labs. Here the parts to be identified are *endmembers* or spectral signatures of some the material that make up space objects, such as aluminum, mylar, solar cell and white paint. Our methods were also used to analyze actual experimental spectral data obtained using the Spica spectrometer, located in the rear blanchard of the 1.6 meter telescope at the Air Force Maui Optical and Supercomputing Site (AMOS) atop Haleakala, Maui. A novel aspect of our approach is the joint identification of endmembers and computation of corresponding fractional abundances. Results of this work were presented at the 2004 AMOS Technical Conference in both oral and poster forms. This work is currently being extended by the GRA Jonathan Piper in his Master's thesis, which is to be completed by May, 2005. A total of four papers [1,4,5,6] were published or accepted for publication during the reporting period as a result of our work with constrained nonnegative matrix factorizations.

Significant collaboration was carried out on the above research topics with members of the Air Force sponsored PRET project, including Robert Plemmons and Todd Torgersen from Wake Forest University, Sudhakar Prasad from the University of New Mexico, and Maile Giffin from Oceanit Labs. Additional collaboration included work with Kris Hamada from Boeing LTS Inc. on the SOI problem, and Michael Berry from the University of Tennessee at Knoxville on text mining applications (document clustering). Finally, the PI in conjunction with Robert Plemmons expect future potential and fruitful collaboration with Kira Jorgensen from the NASA Johnson Space Center on SOI related problems, and with Stuart Jeffries on other PRET related research. The next two subsections present a more detail summary of our accomplishments.

1.1 Design of integrated optical-digital imaging systems for extended depth of field

In a seminal paper, Dowski and Cathey proposed an integrated optical-digital imaging modality, known as wavefront coding, for extending the depth of field (the portion of an image that is in focus). In this approach the basic idea is the use of a cubic phase mask (or lens) in the pupil of a standard, limited-focus imaging system to encode an image of a three-dimensional object than can then be digitally restored (see Figure 1). The surface of the cubic phase mask is given by a function of the form, $\phi(\mathbf{x}) = \alpha(x^3 + y^3)$, where α determines the strength of the mask. Under suitable conditions, restored phase-encoded images exhibit depth-dependent detail that, without any phase encoding, would be washed out because of the normal focus-dependent variant blur.

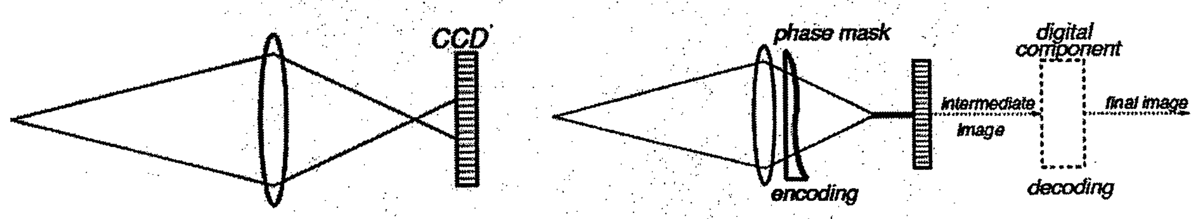


Figure 1: Diagrams of conventional (left) and integrated optical-digital (right) systems.

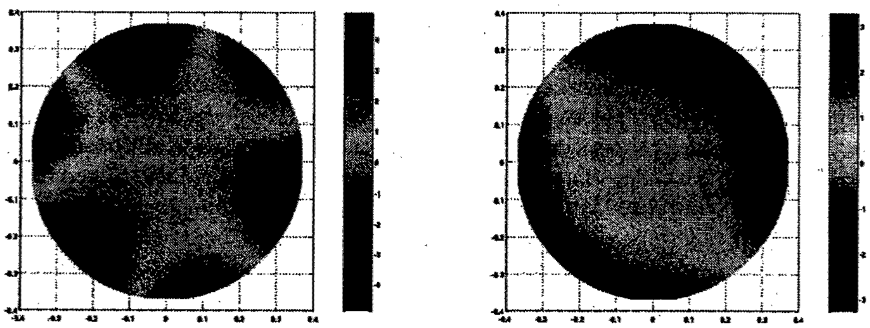
Recently, Prasad and collaborators proposed the concept of pupil phase engineering (PPE), a procedure to optimize the design of a more general pupil-phase mask that can lead to even better performance than the cubic mask in extending the depth of focus and controlling focus-related aberrations. The fundamental basis of this procedure is to seek a phase mask that allows the greatest possible insensitivity of the phase-encoded optical image to focus variation without unacceptably

compromising the digital restorability of that image. Two approaches for optimizing the design of such phase masks were considered: 1) maximizing the insensitivity of the Strehl ratio to focus errors as measured by the smallness of its low-order derivatives and 2) minimizing a Fisher information (FI) norm that determines the integrated sensitivity of the full PSF to focus errors over a range of focal depths.

A third optimization approach, developed by the PI, was proposed and investigated in this project (publications [7,8,9]). In this approach one seeks to maximize the restorability of the intermediate phase-encoded image, as measured by the sum of the singular values of the image blurring matrix, over the sought range of focal depths. Restorability (sum of singular values) is increased when a phase mask boosts information in the signal well above the noise level. We proposed and studied two design optimization metrics:

1. **Penalized FI approach**: a penalized form of the Fisher information norm [2,3] for joint minimization of insensitivity to defocus and maximization of restorability, and

2. **Minimax approach**: a minimization-maximization problem for maximizing the minimum of the sum of singular values over a prescribed defocus range.



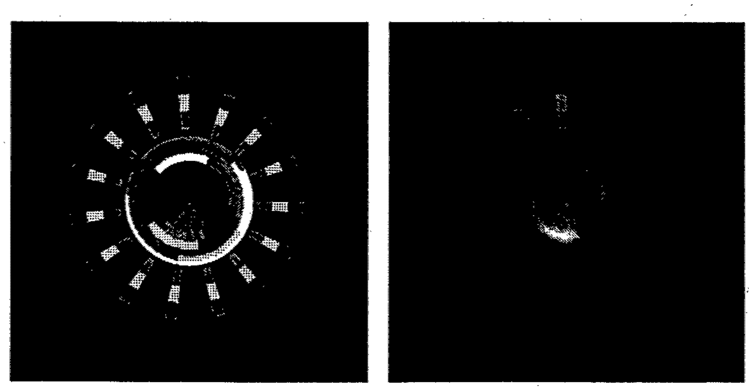
Penalized FI approach Minimax approach

Figure 2: Top-down view of the surface of the optimized phase mask. Red color denotes the peaks and blue denotes the valleys.

The overall effect on digital algorithms includes faster convergence by iterative algorithms and, more importantly, much higher quality than what can be obtained through digital image restoration alone. The classes of phase masks investigated had the following general form,

$$\theta(x,y) = a_1(x^3+y^3) + a_2(x^2y+y^2) + a_3(x^5+y^5) + a_4(x^4y+y^4x) + a_5(x^3y^2+y^3x^2) + \cdots . \tag{1}$$

Extensive numerical computation was conducted to solve these problems using Levenberg-Marquardt as well as minimimax methods for nonlinear optimization. Representatives of the resulting phase mask designs are illustrated in Figure 2.



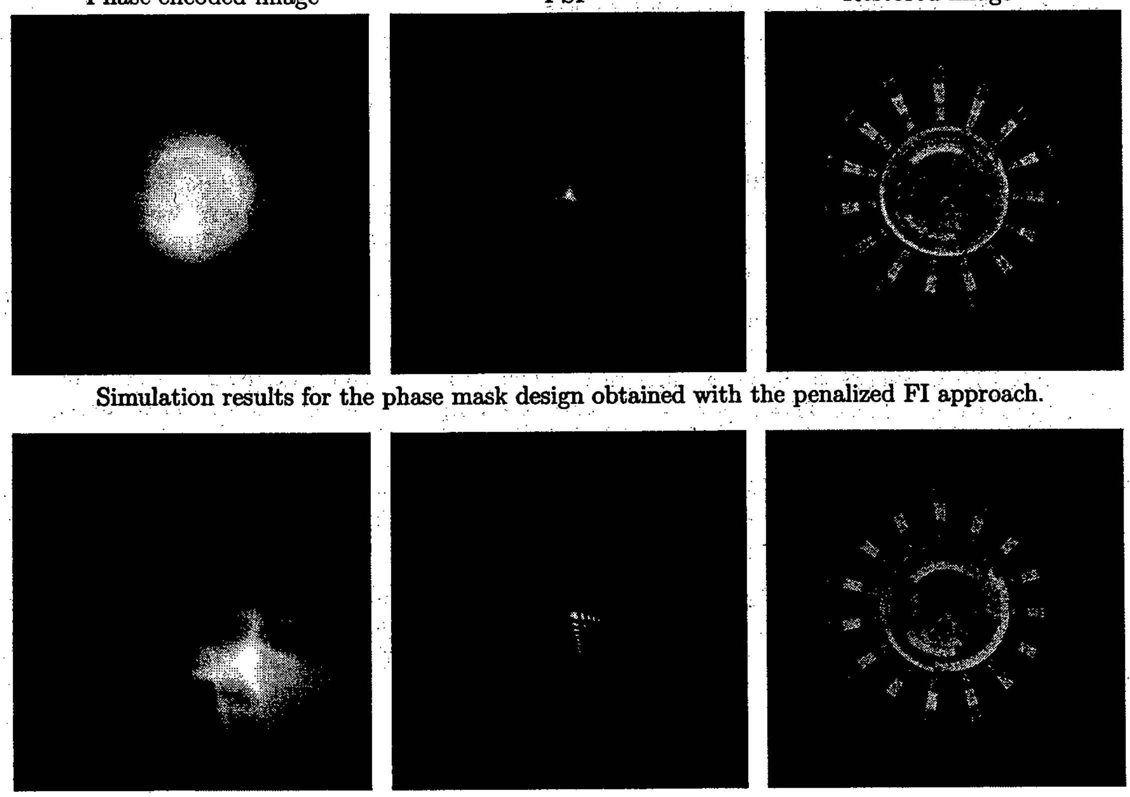
Blur-free perspective projection Spatially varying defocus

Figure 3: Simulated Parabolic Cone Shaped Object.

These phase mask designs were also extensively tested using computer simulation software systems developed by the PI and collaborators. One of these simulators models a parabolic cone-shaped object surrounded by a set of wings arranged in

a spiral at equal angles and linearly progressive distances away from a reference plane coincident with the center point (see Figure 3). Each point in the object is blurred by a spatially varying PSF which models the appropriate defocus according to the distance from that point to the reference plane. Figure 3 illustrates the effect of noise-free spatially variant defocus. For comparison, a theoretical blur-free perspective projection of the object is also shown on the left.

The simulation software integrates the phase mask design into the blurring process. The resulting intermediate and noisy (SNR=100) phase encoded images, the point spread functions, and the final restored images for both phase mask designs are illustrated in Figure 4. Notice the excellent spatial invariance to defocus exhibited by the intermediate images compared to the original spatially varying defocus blur of Figure 3. All images are of size 1024 × 1024. The optimization approaches and



Simulation results for the phase mask design obtained with the penalized FI approach.

Simulation results for the phase mask design obtained with the minimax approach.

Figure 4: The phase-encoded blurred image, the PSF, and the restored image for the resulting phase mask designs.

phase mask designs described in this report are currently being evaluated to extend the depth of field in biometric imaging systems, such as personal identification through iris recognition technology. A patent has been filed on this work.

1.2 Constrained optimization for feature extraction from spectral data

The identification and classification of non-imaging space objects, and ultimately the determination of their shape, function, and status, is an important but difficult problem still to be resolved. While ground-based telescopes with adaptive optics technology have been able to produce high-resolution images for a variety of spaced-based objects, current telescope technology is limited for objects in orbits beyond 1,000 km altitude, such as geosynchronous satellites (approx. 36,000 km orbit). An interesting and promising approach to circumvent this limitation is to collect wavelength-resolved spectral reflectance data, rather than spatially-resolved data, which can then be used to determine information such as material composition of an object. Instruments such as the AMOS Spica spectrometer are often used for this purpose.

1.2.1 A Methodology for the characterization of endmembers and fractional abundances

Current work has been focused on the determination of proportional amounts of known materials (*fractional abundances*) from spectral data. In this project, we extended previous work to determine not only fractional abundances, but also the classes of composing materials that make up the object (*endmembers*). In particular we proposed a methodology for the unsupervised identification and selection of endmembers and computation of fractional abundances that exploit novel *constrained* nonnegative matrix factorization algorithms and information theoretic Kullback-Leibler measures.

Nonnegative matrix factorizations (NMF) were originally proposed by Lee and Seung in a seminal paper in 1999 to find a set of basis functions to represent image data, where the basis functions enable the identification and classification of "intrinsic" parts that make up the object being imaged by multiple observations. These methods have been used in a variety of applications for feature extraction and for approximation of high-dimensional data when the data are comprised of nonnegative components (as is the case with spectral data analysis).

The Kullback-Leibler divergence measure is selected over other methods for matching computed endmembers against laboratory-obtained material traces stored in a database (such as that produced by Kira Jorgensen and collaborators). It has several advantages over simpler methods, such as the cosine of the angle between material traces. The Kullback-Leibler divergence is an information, or cross-entropy measure that captures the spectral correlation between traces and facilitates more accurate matching and selection of computed endmembers.

Our proposed methodology can be described with three words: *extract, select,* and *quantify,* and it is summarized as follows:

Proposed Methodology

1. **Extract candidate endmembers from a set of observed spectral traces.** To do this we solve a nonlinear optimization problem for a set of basis vectors or candidate endmembers, enforcing a smoothness constraint on the calculation of the candidate endmembers. Constrained NMF (CNMF) methods developed by the PI and collaborators are employed.

2. **Select set of endmembers from pool of candidate endmembers.** The Kullback-Leibler divergence is employed to compute proximity of the candidate endmembers to a database of laboratory-obtained material spectral traces. We used a dataset of lab traces corresponding to aluminum, mylar, solar cell, and white paint, among others, provided by Maile Giffin from Oceanit Labs. This dataset contains traces obtained by various researchers at various research labs. Several of these were obtained by Kira Jorgensen from NASA.

3. **Quantify the fractional abundances.** We solve a least squares optimization problem with nonnegativity constraint to quantify the fractional abundances using the selected endmembers. Novel preconditioned iterative methods were employed.

The accuracy of the proposed methodology in the extraction of endmembers and the determination of fractional abundances was extensively tested using simulated spectral data. This simulated data, unlike that used in previous SOI simulation work by other authors, contained additive Gaussian noise and consisted of four test datasets containing 100 spectral traces each. These test datasets corresponded to four different satellites simulated to have varying compositions (fractional abundances) of aluminum, mylar, solar cell, and white paint. The satellites differed from one another by the type of material

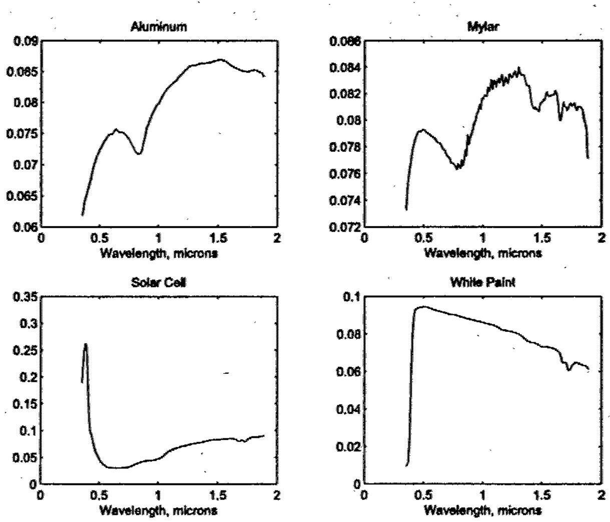
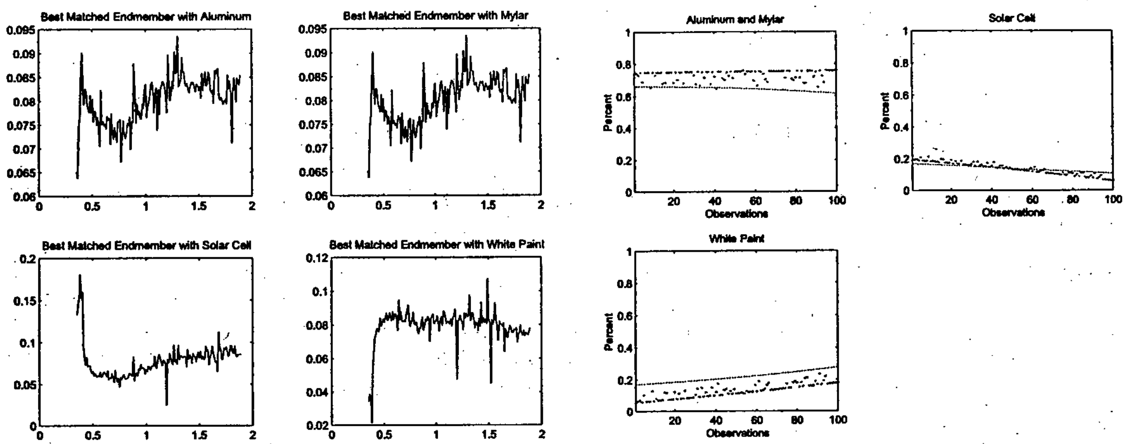


Figure 5: Laboratory spectral traces for the four materials used in data construction: aluminum, mylar, solar cell, and white paint. These spectral traces are in fact averages of several traces of each type of material.

that was deemed to be dominant (approximately 50% fractional abundance for the dominant material, with nearly equal amounts for the remaining three materials). Laboratory-obtained spectral traces for these materials (shown in Figure 5) were employed as the true (unknown) endmembers. The fractional abundances were then allowed to vary sinusoidally with

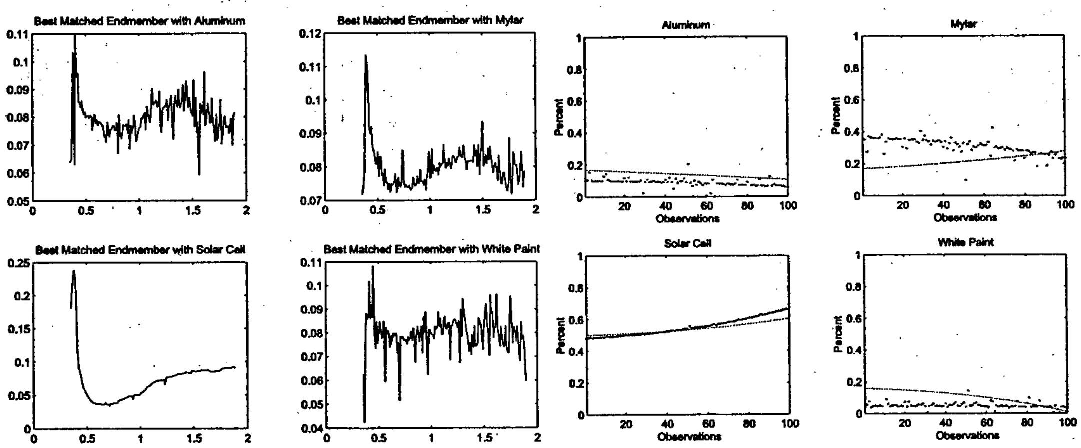
different randomly generated amplitudes, frequencies, and phase shifts, so as to model satellite rotation with respect to a fixed detector, such as the Spica spectrometer.

For each test dataset (different satellite), CNMF was employed to compute a set of candidate endmembers without prior knowledge of the true endmembers or fractional amounts. The Kullback-Leibler divergence measure was then used on the resulting set of candidate endmembers to select the endmembers that best match each of the four original material shown in Figure 5. Finally, given the selected endmembers, preconditioned iterative methods were employed to compute the corresponding fractional abundances. Figure 6 shows the selected endmembers and corresponding fractional abundances for two of the four satellites. Notice the strong similarity between the selected endmembers and the true endmembers shown in Figure 5. More importantly notice the strong correlation between the computed and true fractional abundances as illustrated by the blue and red curves, respectively.



(a) Satellite composed of 50% Aluminum, 17% Mylar, 17% Solar Cell, 16% White Paint.

(b) Simulated satellite composed of 50% Aluminum, 17% Mylar, 17% Solar Cell, 16% White Paint.



(c) Satellite composed of 17% Aluminum, 16% Mylar, 50% Solar Cell, 17% White Paint.

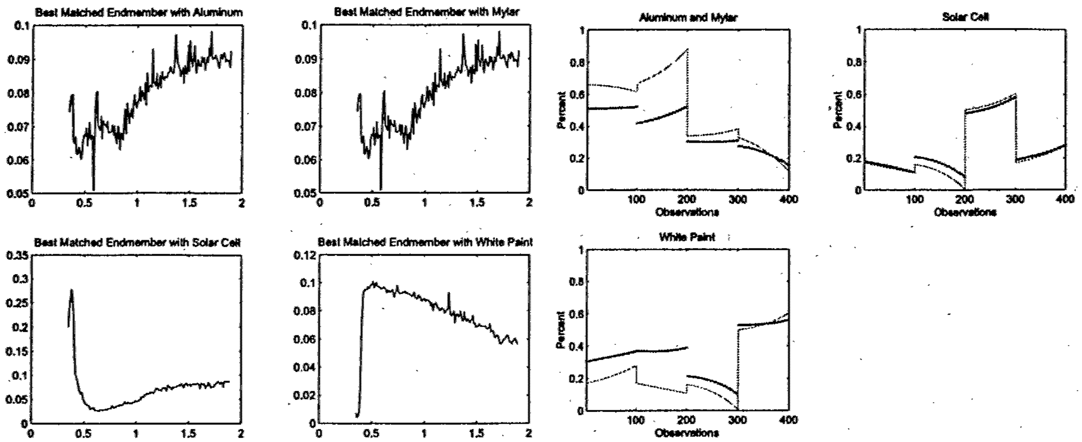
(d) Simulated satellite composed of 17% Aluminum, 16% Mylar, 50% Solar Cell, 17% White Paint.

Figure 6: Selected endmembers (left) and computed fractional abundances (right) for two of the test datasets.

A fifth test dataset was constructed as the composition of the four test datasets in order to increase variability in the data, and analyzed with our approach. The selected endmembers and corresponding fractional abundances are shown in Figure 7. The results of this experiment show even better correlation in the fractional abundances and much better accuracy in the determination of endmembers. The results of this investigation were presented at the 2004 AMOS Technical Conference [5].

1.2.2 Analysis of experimental Spica data

The methods developed in this project were also extensively tested using experimental spectral data collected using the Spica spectrometer located on the rear blanchard of the 1.6 meter telescope at the Air Force Maui Optical and Supercomputing Site (AMOS) atop Haleakala, Maui. In its current configuration the sensor provides 3-4 angstrom resolution spectra in two different modes, a blue mode, which spans the color regime from 4000 angstroms to 7000 angstroms, and a red mode, which



(a) All four simulated satellites. (b) All four simulated satellites.

Figure 7: Selected endmembers (left) and computed fractional abundances (right) for the test dataset consisting of the composition of the four test datasets.

collects 6000 to 9000 angstrom spectra. The spectrometer collects data through a 14 arcsecond diameter aperture, which allows all of the light from an object to enter the instrument and therefore provides absolute photometric information at the expense of slightly limiting the spectral resolution. The spectrometer CCD is a Princeton instruments 1100 × 330 liquid nitrogen cooled array with 25-micron pixels.

For our analysis, a dataset of 1196 different spectral traces pertaining to 22 different space objects was considered. The data was collected over a period of several months. Each trace is stored in an individual data file and contains about 1100 wavelength-intensity pairs. All traces in the dataset were previously corrected for cosmic rays and had background subtractions and the effects of atmospheric absorption removed.

Various NMF methods developed in this project were used for dimensionality reduction of the original dataset and for the extraction of possible material spectra. It is in this latter aspect that our methods excelled compared to other methods, such as principal component analysis. Some of the candidate endmembers found in the dataset are shown in Figure 8. Notice the strong correlation between the candidate endmembers and white paint and solar cell materials. The right most spectra could be identified with mylar.

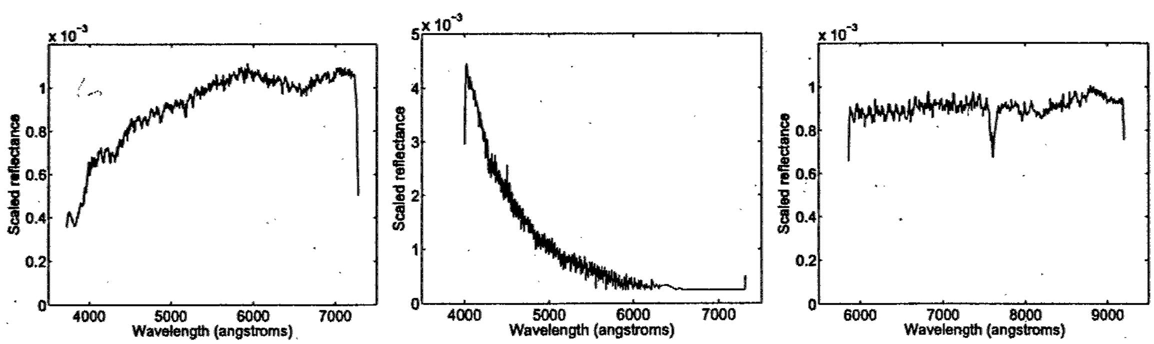


Figure 8: Selected candidate endmembers for the experimental Spica dataset showing intrinsic material spectra for white paint (left) and solar cell (right). The feature in the rightmost position could be identified with mylar.

On a second experiment, a subset of the Spica spectral traces pertaining to a Galaxy V satellite were analyzed with our methods. The resulting candidate endmembers were visually identified with material known to pertain to the Galaxy V satellite by Kira Jorgensen. These results were reported in a poster at the 2004 AMOS Technical Conference [6].

2 Scientific personnel supported by this project

- **Principal Investigator**
  Victor Paul Pauca, Assistant Professor, Department of Computer Science, Wake Forest University

- **Graduate Research Assistant**
  Jonathan Piper, 2nd Year M.S. Graduate Student, Department of Computer Science, Wake Forest University

3 List of papers submitted or published during this report period: February 1, 2003 - June 30, 2004

a. **Papers submitted to peer-reviewed journals**

[1] F. Shahnaz, M. Berry, V. P. Pauca, and R. J. Plemmons. "Document Clustering using Nonnegative Matrix Factorization", submitted to the *Journal on Information Processing & Management*, August 2004. (accepted October 2004)

b. **Papers published in peer-reviewed journals**

[2] S. Prasad, T. C. Torgersen, V. P. Pauca, R. J. Plemmons, and J. van der Gracht. "High-Resolution Imaging Using Integrated Optical Systems", *International Journal on Imaging Systems and Technology*, Special Issue: High-Resolution Image Reconstruction I, Vol. 14, No. 2, pp. 67-74, 2004. (invited submission)

[3] S. Prasad, T. C. Torgersen, V. P. Pauca, R. J. Plemmons and J. van der Gracht. "Deconvolving Space Variant Blue via Pupil-Phase Engineering", invited submission in *Optics in Information Systems*, Special Issue on Computational Imaging, SPIE Int. Tech. Group Newsletter, Vol. 14, No. 2, pp. 4-5, Aug. 2003. (invited submission)

c. **Papers published in peer-reviewed conference proceedings**

[4] V. P. Pauca, F. Shahnaz, M. Berry, and R. J. Plemmons. "Text Mining Using Nonnegative Matrix Factorizations", *Proc. SIAM Int. Conf. on Data Mining*, Orlando, April, 2004.

d. **Papers published or to appear in conference proceedings**

[5] J. Piper, V. P. Pauca, R. J. Plemmons, and M. Giffin. "Object Characterization from Spectral Data Using Nonnegative Matrix Factorization and Information Theory", *Proc. AMOS Tech. Conf.*, Maui, HI, 2004.

[6] V. P. Pauca, R. J. Plemmons, M. Giffin, and K. Hamada. "Mining Scientific Data for Nonimaging Identification and Classification of Space Objects", *Proc. AMOS Tech. Conf.*, Maui, HI, 2004.

[7] S. Prasad, T. Torgersen, P. Pauca, R. Plemmons, and J. van der Gracht. "Pupil-Phase Optimization for Extended-Focus, Aberration-Corrected Imaging Systems", *Proc. SPIE, Advanced Signal Processing: Algorithms, Architectures and Implementation XIV*, Denver, CO, 2004. (to appear)

[8] V. P. Pauca, R. J. Plemmons, S. Prasad, T. C. Torgersen, J. van der Gracht, and C. Vogel. "An Integrated Optical-Digital Approach for Improved Image Restoration", *Proc. AMOS Tech. Conf.* Maui, HI, 2003.

[9] V. P. Pauca, R. J. Plemmons, S. Prasad, T. C. Torgersen, and J. van der Gracht. "Integrated Optical-Digital Approaches for Enhancing Image Restoration and Focus Invariance", *Proc. SPIE, Advanced Signal Processing: Algorithms, Architectures and Implementation XIII*, San Diego, CA, Vol. 5205, pp. 348-357, 2003.

4 Interactions/Transitions

4.1 Participations/Presentations at meetings and conferences

- "Object Characterization from Spectral Data Using Nonnegative Matrix Factorization and Information Theory," oral presentation, 2004 AMOS Technical Conference, Maui, HI (September 2004).
- "Mining Scientific Data for Nonimaging Identification and Classification of Space Objects," poster presentation, 2004 AMOS Technical Conference, Maui, HI (September, 2004).
- "Finding Hidden Components in Spectral Data Using Nonnegative Matrix Factorizations," invited oral presentation, Mathematics & Computing Technology, Boing Corp., Seattle, WA (March 2004).

- "Non-Imaging Identification and Classification of Space Objects from Spectral Sensor Data,", poster presentation by Jonathan Piper, 4th Annual Fitzpatrick Center Annual Meeting–The Physical Nature of Information, Johnathan Piper, Durham, NC (May 2004) (won best student poster award)
- "An Integrated Optical-Digital Approach for Improved Image Restoration," oral presentation, 2003 AMOS Technical Conference, Maui, HI (September 2003).
- "Integrated Optical-Digital Approaches for Enhancing Image Restoration and Focus Invariance," 2004 SPIE Annual Meeting– Advanced Signal Processing: Algorithms, Architectures and Implementation XIII, San Diego, CA (August 2003).

4.2 Transitions/Technology transfers

Significant collaboration and technology transfer was carried out with Maile Giffin from Oceanit Labs and Kris Hamada from Boeing LTS, Inc. Other technology transfer include distribution of our algorithms, both in Matlab and in C, to Michael Berry from the University of Tennessee at Knoxville for text data mining applications.

5 Report of inventions

Inventions are reported on the attached DD882 form.